

---

# Fast Solutions to Projective Monotone Linear Complementarity Problems

---

Geoffrey J. Gordon\*  
Machine Learning Department  
Carnegie Mellon University  
ggordon@cs.cmu.edu

January 1, 2013

## Abstract

We present a new interior-point potential-reduction algorithm for solving monotone linear complementarity problems (LCPs) that have a particular special structure: their matrix  $M \in \mathbb{R}^{n \times n}$  can be decomposed as  $M = \Phi U + \Pi_0$ , where the rank of  $\Phi$  is  $k < n$ , and  $\Pi_0$  denotes Euclidean projection onto the nullspace of  $\Phi^\top$ . We call such LCPs *projective*. Our algorithm solves a monotone projective LCP to relative accuracy  $\epsilon$  in  $O(\sqrt{n} \ln(1/\epsilon))$  iterations, with each iteration requiring  $O(nk^2)$  flops. This complexity compares favorably with interior-point algorithms for general monotone LCPs: these algorithms also require  $O(\sqrt{n} \ln(1/\epsilon))$  iterations, but each iteration needs to solve an  $n \times n$  system of linear equations, a much higher cost than our algorithm when  $k \ll n$ . Our algorithm works even though the solution to a projective LCP is not restricted to lie in any low-rank subspace.

## 1 Linear complementarity problems

The LCP for a matrix  $M \in \mathbb{R}^{n \times n}$  and a vector  $q \in \mathbb{R}^n$  is to find vectors  $x, y \in \mathbb{R}^n$  with

$$x \geq 0 \quad y \geq 0 \quad y = Mx + q \quad x^\top y = 0 \quad (1)$$

We say that vectors  $x, y$  are *feasible* if they satisfy the first three conditions of (1) (i.e., leaving off complementarity), and we call them a *solution* if they satisfy all four conditions. The complementarity gap  $x^\top y$  is nonnegative for any feasible point  $(x, y)$ , and measures how close a feasible point is to being a solution. (See [1] for an overview of LCPs.)

---

\*This work was first presented at the 2010 NIPS workshop “Learning and Planning from Batch Time Series Data.”

If  $M$  is positive semidefinite (but not necessarily symmetric), the LCP is *monotone*, and there exist interior-point algorithms that solve it to relative accuracy  $\epsilon$  in  $O(\sqrt{n} \ln(1/\epsilon))$  Newton-like iterations. In each iteration, the main work is to solve an  $n \times n$  system of linear equations.

Suppose the matrix  $M$  can be decomposed as  $M = \Phi U + \Pi_0$ , where  $\Phi \in \mathbb{R}^{n \times k}$  and  $U \in \mathbb{R}^{k \times n}$  have rank  $k < n$ , and  $\Pi_0$  projects onto the nullspace of  $\Phi^\top$  (that is,  $\Pi_0 = I - \Phi \Phi^\dagger$ , where  $\dagger$  denotes the Moore-Penrose pseudoinverse). In this case we call  $(M, q)$  a *projective* LCP of rank  $k$ . Our new algorithm solves a projective LCP in  $O(\sqrt{n} \ln(1/\epsilon))$  iterations, the same as for the general monotone case, but with each iteration requiring only  $O(nk^2)$  flops.

This result is an analog of the situation for linear equations: a rank- $k$  factored system of linear equations can be solved in  $O(nk^2)$  flops, while it is believed that a general  $n \times n$  system of equations requires  $\Omega(n^{2+\eta})$  flops for some constant  $\eta > 0$ . However, unlike the situation for linear equations, in a projective LCP we can’t *a priori* restrict either  $x$  or  $y$  to a low-rank subspace of  $\mathbb{R}^n$ ; so, it is perhaps surprising that the analogous complexity result still holds. (The inequality constraints  $x \geq 0$  and  $y \geq 0$  are the source of this difficulty: the intersection of  $x \geq 0$  or  $y \geq 0$  with a rank- $k$  subspace can be quite restrictive.)

## 2 Potential reduction

We say that  $x$  and  $y$  are *strictly feasible* if they satisfy  $x > 0$ ,  $y > 0$ , and  $y = Mx + q$ . We will assume that we know a strictly feasible initial point  $(x_0, y_0)$  for our LCP. (If we do not, it is possible to construct one, as mentioned in [2].)

For any strictly feasible point  $(x, y)$ , fixing a parameter  $\kappa > 0$ , we define the potential

$$p_\kappa(x, y) = (n + \kappa) \ln x^\top y - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln y_i \quad (2)$$

We will design an algorithm that attempts to reduce the potential  $p_\kappa(x, y)$  over time. The following lemma justifies this idea:

**Lemma 2.1** *For any strictly feasible  $x, y$ ,*

$$p_0(x, y) = n \ln x^\top y - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln y_i > 0 \quad (3)$$

(For a proof, see the appendix.) In particular, Lemma 2.1 implies  $p_\kappa(x, y) \geq \kappa \ln x^\top y$ , or  $x^\top y \leq \exp(p_\kappa(x, y)/\kappa)$ ; so, if we can reduce the potential by at least some amount  $\delta > 0$  per iteration, after  $T$  iterations our iterate  $(x, y)$  will satisfy

$$x^\top y \leq \exp((p_\kappa(x_0, y_0) - T\delta)/\kappa)$$

That is, our algorithm will converge linearly: a bound on the gap will decrease by a factor of  $\exp(-\delta/\kappa)$  per iteration. Below, we will take  $\kappa = \sqrt{n}$ , and  $\delta$  will not depend on  $n$ ; so, if we desire a reduction of our potential by a factor  $0 < \epsilon < 1$ , we will need  $\frac{\sqrt{n}}{\delta} \ln(1/\epsilon)$  iterations, as the abstract states.

### 3 The central path

Lemma 2.1 shows that  $p_0(x, y) > 0$ . The local minimizers of  $p_0$  are the points where its gradient vanishes:

$$\begin{aligned} 0 &= \frac{d}{dx_i} p_0(x, y) = ny_i/x^\top y - 1/x_i \\ 0 &= \frac{d}{dy_i} p_0(x, y) = nx_i/x^\top y - 1/y_i \end{aligned}$$

Multiplying the first equation through by  $x_i(x^\top y)/n$ , or the second equation through by  $y_i(x^\top y)/n$ , we get

$$x_i y_i = x^\top y / n$$

which is satisfied for a pair  $x, y > 0$  if and only if  $x_i y_i = t$  for all  $i$  and some  $t > 0$ . Equivalently, we can write  $x \circ y = t\mathbf{1}$ , where  $\circ$  denotes the Hadamard (elementwise) product.

The points  $(x, y)$  that satisfy

$$x > 0 \quad y > 0 \quad y = Mx + q \quad x \circ y = t\mathbf{1}$$

are called the *central path* of the LCP  $(M, q)$ ; for monotone LCPs, if the central path is nonempty, it is a smooth curve, and it approaches the solution of the

LCP as  $t \rightarrow 0$ . We can view the term  $p_0(x, y)$  as encouraging our algorithm to remain close to the central path; the remaining part of our potential,  $\kappa \ln(x^\top y)$ , encourages our algorithm to slide along the central path, reducing  $t = x^\top y/n$  and pushing us closer to a solution.

### 4 A result on rank

To make our algorithm run in time  $O(nk^2)$  per iteration, we will need to do most of our calculations on vectors of length  $k$  instead of length  $n$ . Unfortunately, as mentioned earlier, the vectors  $x$  and  $y$  are not guaranteed to lie in any rank- $k$  subspace. Our main insight is that we can work mostly from a function of  $x$  and  $y$  that does lie in a rank- $k$  subspace. In more detail:

**Lemma 4.1** *Suppose the pair  $x, y$  is feasible for the monotone LCP  $(M, q)$ , and that  $M = \Phi U + \Pi_0$ , where  $\Phi$  has rank  $k$  and  $\Pi_0$  projects onto the nullspace of  $\Phi^\top$ . Then the vector  $x - y + q$  is in the range of  $\Phi$ .*

PROOF: Define  $\Pi = I - \Pi_0$ , so that the range of  $\Pi$  is the same as the range of  $\Phi$ . Since  $\Pi_0$  projects onto the nullspace of  $\Phi^\top$ , we know  $\Pi_0 \Phi = 0$ . And, since  $\Pi_0$  is a projection matrix, we have  $\Pi_0^2 = \Pi_0$ . So,  $\Pi_0 M = \Pi_0(\Phi U + \Pi_0) = \Pi_0$ . Therefore, for any feasible  $x, y$ :

$$\begin{aligned} y &= Mx + q \\ \Pi_0 y &= \Pi_0 Mx + \Pi_0 q \\ \Pi_0 y &= \Pi_0 x + \Pi_0 q \\ (I - \Pi)y &= (I - \Pi)x + (I - \Pi)q \\ \Pi(x - y + q) &= x - y + q \end{aligned}$$

So,  $x - y + q$  is in the range of  $\Pi$ , as claimed.  $\square$

In general, we can't recover  $x$  or  $y$  individually from  $x - y + q$ . However, if we know that the pair  $x, y$  solves the LCP, we can use complementarity to recover  $x$  and  $y$ :  $x^\top y = 0$ , so for any  $i$ , at most one of  $x_i$  and  $y_i$  can be nonzero. So, given  $x - y + q$ , we subtract  $q$  to get  $z = x - y$ . Then we set  $x = z_+$  and  $y = z_-$ , i.e.,  $x_i = \max(z_i, 0)$  and  $y_i = \max(-z_i, 0)$ .

At intermediate points in our algorithm, we maintain  $x$  and  $y$  separately, and constrain  $x - y + q = \Phi w$ . We calculate the update for  $w$  first by manipulating length- $k$  vectors, and then use this result to derive updates for  $x$  and  $y$  with work that is only linear in  $n$ .

### 5 The algorithm

We will base our algorithm on a potential-reduction method due to Kojima et al. [2]. The algorithm

---

In: LCP  $(M, q)$ ; strictly feasible  $x, y$ ;  $\beta \in (0, 1)$ ;  $\epsilon > 0$ .  
 Out: strictly feasible  $x, y$  with  $x^\top y \leq \epsilon$ .

---

1. Stop if  $x^\top y \leq \epsilon$ .
  2. Solve (4-5) for an update  $(\Delta x, \Delta y)$ .
  3. Choose a step length  $\theta \geq 0$  by (7).
  4. Set  $(x, y) \leftarrow (x, y) + \theta(\Delta x, \Delta y)$ .
  5. Repeat from step 1.
- 

Figure 1: Potential-reduction algorithm.

(Fig. 1) uses Newton's method to step toward a point on the central path with  $x \circ y = t\mathbf{1}$ . It has a single parameter  $\beta \in (0, 1)$  that helps us choose the target complementarity  $t$ : we set  $t$  to be a fraction  $\beta$  of the current average complementarity, i.e.,  $t = \beta \frac{x^\top y}{n}$ . Taking  $\beta$  near 1 causes us to follow the central path closely, reducing  $t$  slowly; taking  $\beta$  near 0 tries aggressively to reduce the complementarity gap, but risks straying farther from the central path.<sup>1</sup>

We will derive the algorithm first for general  $M$  (in which case it is the same as that of Kojima et al.), and then specialize it to the projective case  $M = \Phi U + \Pi_0$ . We will assume that  $M$  is positive semidefinite, i.e., that  $\Phi U$  is positive semidefinite.

To step toward a point on the central path with complementarity  $t$ , we want to satisfy both  $x \circ y = t\mathbf{1}$  and  $y = Mx + q$ . The first-order Taylor approximation to  $x \circ y = t\mathbf{1}$  is

$$x \circ y + \Delta x \circ y + x \circ \Delta y = t\mathbf{1}$$

Writing  $X = \text{diag}(x)$  and  $Y = \text{diag}(y)$ , the Newton step is therefore

$$\begin{pmatrix} Y & X \\ -M & I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} g \\ r \end{pmatrix} \quad (4)$$

where

$$\begin{aligned} g &= \beta \frac{x^\top y}{n} \mathbf{1} - x \circ y \\ r &= Mx + q - y \end{aligned} \quad (5)$$

Here we have used  $t = \beta \frac{x^\top y}{n}$ . Note also that  $r = 0$  for any feasible  $x, y$ .

---

<sup>1</sup>For simplicity of notation we require  $x \geq 0$  and include only inequality constraints, but the algorithm works with minor changes if we allow some components of  $x$  to be free instead of nonnegative, in which case the corresponding constraints are equalities instead of inequalities (i.e., the corresponding elements of  $y$  must be zero).

To pick a step length, we first define a vector  $s$  and diagonal matrix  $S$  with

$$s_i = S_{ii} = \sqrt{x_i y_i} \quad (6)$$

Write  $s_0$  for the smallest element of  $s$ . Like  $g$ , the vector  $s/s_0$  measures how far we are from the central path: if we are near the central path,  $s/s_0 \approx \mathbf{1}$ , and we can afford to take a relatively larger step, while if we are far from the central path, some elements of  $s/s_0$  will be large, and we will need to be more cautious.

In particular, we will show below that the step size

$$\theta = \frac{3}{7} \frac{s_0}{\|S^{-1}g\|} \quad (7)$$

guarantees that we maintain strict feasibility and decrease our potential. Note that  $s > 0$  for any strictly feasible  $x, y$ ; and, since  $\beta < 1$ , we have  $g \neq 0$  for any strictly feasible  $x, y$ . So, (7) always yields a well-defined step length  $\theta$ . In practice, the value of  $\theta$  from (7) will be conservative; but, it could serve as an initializer for a line search (e.g., [3, Alg. 9.2]) to determine a step length that decreases the potential as much as possible.

## 6 Proof of correctness

We proceed to show that the potential-reduction algorithm behaves as claimed above. The proof follows Kojima et al. [2], although our presentation is somewhat different.

**Lemma 6.1** *Eq. 4 defines a unique step direction.*

For a proof, see the appendix.

**Theorem 6.1** *Suppose  $n \geq 2$ , and take the parameter  $\kappa$  from our potential (2) to be  $\kappa = \sqrt{n}$ . Suppose  $x$  and  $y$  are strictly feasible for the monotone LCP  $(M, q)$ . Then the step  $(\theta\Delta x, \theta\Delta y)$  from (4-7), with  $\beta = \frac{n}{n+\kappa}$ , maintains strict feasibility and guarantees a potential reduction of at least:*

$$p_\kappa(x + \theta\Delta x, y + \theta\Delta y) \leq p_\kappa(x, y) - \frac{1}{5}$$

**PROOF (THM. 6.1):** To make notation simpler, we will change variables to  $u = X^{-1}\Delta x$  and  $v = Y^{-1}\Delta y$ . With this notation, and using (6), the first row of (4) becomes

$$S^2(u + v) = g \quad (8)$$

Our goal is now to bound the change in potential

$$\Delta p = p_\kappa(x + \theta\Delta x, y + \theta\Delta y) - p_\kappa(x, y)$$

We start by splitting  $p_\kappa(x, y)$  into two pieces, so that we can bound each piece separately:

$$\begin{aligned} p_\kappa(x, y) &= \bar{p}_1(x, y) + \bar{p}_2(x, y) \\ \bar{p}_1(x, y) &= \frac{n}{\beta} \ln x^\top y \\ \bar{p}_2(x, y) &= -\sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln y_i \end{aligned}$$

Here we have used  $n + \kappa = \frac{n}{\beta}$ , from the assumed value of  $\beta$ . Note that  $\bar{p}_2$  is convex, but  $\bar{p}_1$  is not (due to the concave  $\ln$  function and the interaction  $x^\top y$ ). Write

$$\begin{aligned} \Delta \bar{p}_1 &= \bar{p}_1(x + \theta \Delta x, y + \theta \Delta y) - \bar{p}_1(x, y) \\ \Delta \bar{p}_2 &= \bar{p}_2(x + \theta \Delta x, y + \theta \Delta y) - \bar{p}_2(x, y) \end{aligned}$$

To upper bound  $\Delta \bar{p}_1$  we will use the identity

$$\ln(z + \Delta z) \leq \ln z + \Delta z / z$$

which holds since  $\ln(z)$  is concave in  $z$ . We take  $z = x^\top y$ , so

$$\begin{aligned} \Delta z &= (x + \theta \Delta x)^\top (y + \theta \Delta y) - x^\top y \\ &= \theta \Delta x^\top y + \theta x^\top \Delta y + \theta^2 \Delta x^\top \Delta y \end{aligned}$$

Therefore,

$$\Delta \bar{p}_1 \leq \frac{n}{\beta x^\top y} [\theta \Delta x^\top y + \theta x^\top \Delta y + \theta^2 \Delta x^\top \Delta y]$$

or, in terms of  $u$  and  $v$ ,

$$\Delta \bar{p}_1 \leq \frac{n}{\beta x^\top y} [\theta (x \circ y)^\top (u + v) + \theta^2 \gamma]$$

where we have written  $\gamma = \Delta x^\top \Delta y = u^\top X Y v$ . Note that  $\gamma \geq 0$ , since  $\gamma = \Delta x^\top \Delta y = \Delta x^\top M \Delta x$  and  $M$  is positive semidefinite.

For  $\Delta \bar{p}_2$  we use a local upper bound on  $-\ln(z + \Delta z)$ , derived from the second-order Taylor approximation

$$-\ln(z + \Delta z) \approx -\ln z - \Delta z / z + \frac{1}{2} (\Delta z)^2 / z^2$$

To get a bound valid for some range of  $\Delta z$ , we scale up the second derivative by a factor  $\tau \geq 1$ :

**Lemma 6.2** *For  $\tau \geq 1$ , if  $\frac{\Delta z}{z} \geq \frac{1-\tau}{\tau}$ , then*

$$-\ln(z + \Delta z) \leq -\ln z - \Delta z / z + \frac{\tau}{2} (\Delta z)^2 / z^2 \quad (9)$$

(See the appendix for a proof.) So, using Lemma 6.2  $2n$  times (first with  $z = x_i$  and  $\Delta z = \theta \Delta x_i$ , and then with  $z = y_i$  and  $\Delta z = \theta \Delta y_i$ ), we have

$$\begin{aligned} \Delta \bar{p}_2 &\leq -\theta \sum_{i=1}^n \Delta x_i / x_i - \theta \sum_{i=1}^n \Delta y_i / y_i \\ &\quad + \theta^2 \frac{\tau}{2} \sum_{i=1}^n (\Delta x_i)^2 / x_i^2 + \theta^2 \frac{\tau}{2} \sum_{i=1}^n (\Delta y_i)^2 / y_i^2 \end{aligned}$$

so long as  $\theta \Delta x_i / x_i \geq \frac{1-\tau}{\tau}$  and  $\theta \Delta y_i / y_i \geq \frac{1-\tau}{\tau}$  for all  $i$ . Or, in terms of  $u$  and  $v$ ,

$$\Delta \bar{p}_2 \leq -\theta \mathbf{1}^\top (u + v) + \frac{\tau}{2} \theta^2 (u^\top u + v^\top v)$$

as long as

$$\theta u \geq \frac{1-\tau}{\tau} \mathbf{1} \quad \theta v \geq \frac{1-\tau}{\tau} \mathbf{1} \quad (10)$$

Combining the bounds on  $\Delta \bar{p}_1$  and  $\Delta \bar{p}_2$ , we have

$$\begin{aligned} \Delta p &\leq \frac{n}{\beta x^\top y} [\theta (x \circ y)^\top (u + v) + \theta^2 \gamma] \\ &\quad - \theta \mathbf{1}^\top (u + v) + \frac{\tau}{2} \theta^2 (u^\top u + v^\top v) \end{aligned} \quad (11)$$

as long as (10) holds. Using the definition (5) of  $g$ , we can split the right-hand side of (11) into a term that is linear in  $\theta$ :

$$-\frac{n}{\beta x^\top y} \theta g^\top (u + v)$$

and a term that is quadratic in  $\theta$ :

$$\theta^2 [\frac{n}{\beta x^\top y} \gamma + \frac{\tau}{2} (u^\top u + v^\top v)]$$

We can simplify each of these terms separately: using (8), we have

$$g^\top (u + v) = g^\top S^{-2} g = \|S^{-1} g\|^2$$

And,

$$\begin{aligned} u^\top u + v^\top v &= \|S^{-1} S u\|^2 + \|S^{-1} S v\|^2 \\ &\leq \frac{1}{s_0^2} (\|S u\|^2 + \|S v\|^2) \\ &= \frac{1}{s_0^2} (\|S(u + v)\|^2 - 2\gamma) \\ &= \frac{1}{s_0^2} (\|S^{-1} g\|^2 - 2\gamma) \end{aligned}$$

(The second line holds by definition of  $s_0$ ; the third uses the definition of  $\gamma$ ; and the last uses (8) again.)

So, (11) becomes

$$\begin{aligned} \Delta p &\leq \theta^2 [(\frac{n}{\beta x^\top y} - \frac{\tau}{s_0^2}) \gamma + \frac{\tau}{2 s_0^2} \|S^{-1} g\|^2] \\ &\quad - \theta \frac{n}{\beta x^\top y} \|S^{-1} g\|^2 \end{aligned} \quad (12)$$

as long as (10) holds. In Lemma 6.2, we are free to choose  $\tau \geq 1$ ; so, we will assume

$$\tau \geq \frac{1}{\beta} \quad (13)$$

So, since

$$\frac{x^\top y}{n} = \frac{s^\top s}{n} \geq s_0^2$$

we have that  $(\frac{n}{\beta x^\top y} - \frac{\tau}{s_0^2}) \gamma \leq 0$ , and (12) becomes

$$\Delta p \leq \theta^2 \frac{\tau}{2 s_0^2} \|S^{-1} g\|^2 - \theta \frac{n}{\beta x^\top y} \|S^{-1} g\|^2 \quad (14)$$

as long as (10) and (13) hold. The right-hand side will be negative for the optimal  $\theta > 0$ , since its derivative

with respect to  $\theta$  is negative at  $\theta = 0$ . So, we now know that  $\Delta p < 0$ , i.e.,  $(\Delta x, \Delta y)$  is a descent direction for  $p$  as desired.

To determine how large a decrease in potential we can achieve, we need to pick a feasible step size  $\theta$ . To ensure that we satisfy (10), we will enforce the stricter constraints

$$\|\theta u\|_\infty \leq \frac{\tau-1}{\tau} \quad \|\theta v\|_\infty \leq \frac{\tau-1}{\tau} \quad (15)$$

Note that (15) implies that our step maintains strict feasibility: since  $\frac{\tau-1}{\tau} < 1$ , we have  $\|\theta u\|_\infty = \|\theta X^{-1}\Delta x\|_\infty < 1$  and  $\|\theta v\|_\infty = \|\theta Y^{-1}\Delta y\|_\infty < 1$ . Now,

$$\begin{aligned} \|u\|_\infty &\leq \|u\| \\ &= \|S^{-1}Su\| \\ &\leq \frac{1}{s_0}\|Su\| \\ &\leq \frac{1}{s_0}(\|Su\| + \|Sv\| + 2\gamma) \\ &= \frac{1}{s_0}\|Su + Sv\| \\ &= \frac{1}{s_0}\|S^{-1}g\| \end{aligned}$$

Analogously,

$$\|v\|_\infty \leq \frac{1}{s_0}\|S^{-1}g\|$$

So, (15) will be satisfied if we take

$$\theta = \frac{\tau-1}{\tau} \frac{s_0}{\|S^{-1}g\|} \quad (16)$$

Substituting into (14), we have

$$\begin{aligned} \Delta p &\leq \frac{(\tau-1)^2}{\tau^2} \frac{s_0^2}{\|S^{-1}g\|^2} \frac{\tau}{2s_0^2} \|S^{-1}g\|^2 \\ &\quad - \frac{\tau-1}{\tau} \frac{s_0}{\|S^{-1}g\|} \frac{n}{\beta x^\top y} \|S^{-1}g\|^2 \\ &= \frac{1}{2} \frac{(\tau-1)^2}{\tau} - \frac{\tau-1}{\tau} \frac{ns_0}{\beta s^\top s} \|S^{-1}g\| \end{aligned} \quad (17)$$

as long as (13) holds. Finally, we lower-bound  $\|S^{-1}g\|$  with the following lemma, whose proof is in the appendix:

**Lemma 6.3** *For  $g$  in (5), if  $\beta = \frac{n}{n+\sqrt{n}}$ , then:*

$$\|S^{-1}g\| \geq \frac{\sqrt{3}}{2} \frac{s^\top s}{n} \frac{\beta}{s_0}$$

Substituting Lemma 6.3 into (17), we have

$$\Delta p \leq \frac{1}{2} \frac{(\tau-1)^2}{\tau} - \frac{\tau-1}{\tau} \frac{\sqrt{3}}{2}$$

In particular, if we take  $\tau = \frac{7}{4}$ , we satisfy (13):  $n \geq 2$ , so  $\frac{1}{\beta} \leq \frac{2+\sqrt{2}}{2} \approx 1.707$ . And, we have

$$\Delta p \leq \frac{1}{2} \frac{9}{16} \frac{4}{7} - \frac{3}{7} \frac{\sqrt{3}}{2} \leq -\frac{1}{5}$$

as claimed. This value of  $\tau$ , together with (16), yields (7).  $\square$

---

In:  $\Phi, U, q$ ; strictly feasible  $x, y$ ;  $\beta \in (0, 1)$ ;  $\epsilon > 0$ .  
 Out: strictly feasible  $x, y$  with  $x^\top y \leq \epsilon$ .

---

1. Set  $w \leftarrow \Phi^\dagger(x - y + q)$ .
  2. Stop if  $x^\top y \leq \epsilon$ .
  3. Compute  $g$  and  $r$  from (5) and (24).
  4. Compute  $G$  and  $h$  via (22–23).
  5. Solve  $G\Delta w = h$  for  $\Delta w$ .
  6. Solve  $(X + Y)\Delta y = g - Y\Phi\Delta w$  for  $\Delta y$ .
  7. Compute  $\Delta x = \Delta y + \Phi\Delta w$ .
  8. Choose a step length  $\theta \geq 0$  by (7).
  9. Set  $(x, y, w) \leftarrow (x, y, w) + \theta(\Delta x, \Delta y, \Delta w)$ .
  10. Repeat from step 2.
- 

Figure 2: Potential reduction for projective LCPs.

## 7 Algorithm for projective LCPs

The main work in each iteration of the potential-reduction algorithm is to compute the Newton direction (4), which requires solving an  $n \times n$  system of linear equations. (The system (4) as a whole is  $2n \times 2n$ , but we can use the sparsity of the three diagonal blocks to eliminate cheaply down to an  $n \times n$  system.) The work required to solve this  $n \times n$  system can vary greatly, depending on the structure of  $M$ , but is often prohibitive for large  $n$ .

So, in the projective case ( $M = \Phi U + \Pi_0$  and  $\Pi_0 = I - \Phi\Phi^\dagger$ , where  $\Phi$  has  $k < n$  columns), we want to avoid solving an  $n \times n$  system at all; instead we will construct and solve only a smaller  $k \times k$  system. Constructing the  $k \times k$  system will then be the main work in each iteration, at  $O(nk^2)$  flops. (Solving the  $k \times k$  system takes at most  $O(k^3)$  flops even if we just use simple Gaussian elimination.)

To run our potential-reduction algorithm on a projective LCP, our basic idea (as discussed in Sec. 4) is to keep track of  $w$  such that  $x - y + q = \Phi w$ , and do as many calculations as possible in terms of  $w$  instead of  $x$  and  $y$ . Fig. 2 summarizes the resulting algorithm. (In fact it is not even necessary to keep track of  $w$  explicitly, but Fig. 2 makes  $w$  explicit for clarity.) For convenience we assume that  $\Phi$  has full column rank; if not, we can drop some columns from  $\Phi$  and adjust  $U$  accordingly.

We are given a strictly feasible pair  $x, y$  to start, so our initial  $w$  is just  $\Phi^\dagger(x - y + q)$ ; we then have  $x - y + q = \Phi w$  by Lemma 4.1. We update  $w$  by adjoining the equation  $\Delta x - \Delta y = \Phi\Delta w$  to the system (4). With

this extra constraint, (4) becomes:

$$\begin{pmatrix} Y & X & 0 \\ -M & I & 0 \\ -I & I & \Phi \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta w \end{pmatrix} = \begin{pmatrix} g \\ r \\ 0 \end{pmatrix} \quad (18)$$

Note that the extra constraint does not change the sequence of points  $(x, y)$  visited by our potential reduction algorithm: its only effect is to allow us to track  $w$  and solve (18) efficiently.

To solve (18) efficiently, we will run several steps of block Gaussian elimination analytically. First use the last block row of (18) to eliminate the first block column:

$$\begin{pmatrix} X + Y & Y\Phi \\ I - M & -M\Phi \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta w \end{pmatrix} = \begin{pmatrix} g \\ r \end{pmatrix} \quad (19)$$

Then note that strict feasibility implies that  $X + Y$  is nonsingular. So, we can use the first block row of (19) to eliminate the first block column:

$$[(M - I)(X + Y)^{-1}Y\Phi - M\Phi]\Delta w = (M - I)(X + Y)^{-1}g + r \quad (20)$$

Finally, we can left-multiply (20) by  $\Phi^\top$  to reduce to

$$G\Delta w = h \quad (21)$$

where

$$\begin{aligned} G &= \Phi^\top[(M - I)(X + Y)^{-1}Y - M]\Phi \\ &= (\Phi^\top\Phi U - \Phi^\top)(X + Y)^{-1}Y\Phi - \Phi^\top(\Phi U)\Phi \quad (22) \\ h &= \Phi^\top[(M - I)(X + Y)^{-1}g + r] \\ &= (\Phi^\top\Phi U - \Phi^\top)(X + Y)^{-1}g + \Phi^\top r \quad (23) \end{aligned}$$

Eqs. 22–23 show how to build  $G$  and  $h$  in time  $O(nk^2)$ , starting from  $x, y, r, g, \Phi^\top\Phi U$ , and  $\Phi$ , which together require  $O(nk)$  storage. The vectors  $r$  and  $g$  can be calculated efficiently using (5) and the representation  $M = \Phi U + I - \Phi\Phi^\dagger$ : we compute

$$Mx = \Phi(Ux - \Phi^\dagger x) + x \quad (24)$$

For the term  $\Phi^\dagger x$ , it may help to precompute a factorization such as the  $QR$  decomposition of  $\Phi$ .

We can then solve (21) for  $\Delta w$  in time  $O(k^3)$  or better. (Lemma 7.1, whose proof is in the appendix, ensures that  $\Delta w$  is uniquely determined.) Since  $X + Y$  is nonsingular, we can use the first block row of (19) to solve for  $\Delta y$  in time  $O(nk)$ . Finally, we can use the last block row of (18) to solve for  $\Delta x$  in time  $O(nk)$ .

**Lemma 7.1** *If  $x, y > 0$ ,  $\Phi$  has full column rank, and  $\Phi U$  is positive semidefinite, then the matrix  $G$  defined in (22) is invertible.*

Since  $k < n$ , the total time per iteration is  $O(nk^2)$ , as claimed earlier—potentially substantially faster than an iteration of the potential reduction method on an arbitrary monotone LCP. Since we are performing the exact same sequence of updates to  $x$  and  $y$  as the general potential-reduction algorithm running on  $(M, q)$ , our bounds from Sec. 6 continue to hold: we take the same number of iterations and reach the same final error level. So, we have proven:

**Theorem 7.1** *Suppose  $n \geq 2$ , and take the parameter  $\kappa$  from our potential (2) to be  $\kappa = \sqrt{n}$ . Suppose  $x$  and  $y$  are strictly feasible for the monotone projective LCP  $(M, q)$ , where  $M = \Phi U + \Pi_0$ ,  $\Pi_0 = I - \Phi\Phi^\dagger$ , and  $\Phi$  has full column rank. Then the algorithm of Fig. 2, with  $\beta = \frac{n}{n+\kappa}$ , maintains strict feasibility and guarantees a potential reduction of at least  $\frac{1}{5}$  per step.*

## Acknowledgements

This work was supported by ONR MURI grant number N00014-09-1-1052.

## References

- [1] Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. *The Linear Complementarity Problem*. Academic Press, 1992. Revised edition, SIAM 2009.
- [2] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. An  $O(\sqrt{n}L)$ -iteration potential reduction algorithm for linear complementary problems. *Mathematical Programming*, 50:331–342, 1991.
- [3] Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, 2004.

## A Proofs of Lemmas

PROOF (LEMMA 2.1): Define  $u_i = y_i x_i$  and let  $u = \sum_i u_i$ . Then  $x^\top y = u$ , and

$$\begin{aligned} n \ln x^\top y - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln y_i \\ = - \sum_{i=1}^n \ln u_i / u > 0 \end{aligned}$$

since  $0 < u_i / u < 1$  for all  $i$ .  $\square$

PROOF (LEM. 6.1): Since  $x > 0$ ,  $X$  is invertible. So, we can use Gaussian elimination on (4) to arrive at

$$(-M - X^{-1}Y)\Delta x = Mx + q - y - X^{-1}(t\mathbf{1} - x \circ y)$$

(In particular, subtract  $X^{-1}$  times the first row from the second row).

Since  $M$  is positive semidefinite and  $x, y > 0$ ,  $M + X^{-1}Y$  is strictly positive definite, and so we can solve uniquely for  $\Delta x$ . We can then substitute  $\Delta x$  into the first row of (4), which leads to a unique solution for  $\Delta y$  since  $X$  is invertible.  $\square$

PROOF (LEMMA 6.2): By construction, the left-hand and right-hand sides of (9) match in value and first derivative at  $\Delta z = 0$ . The second derivative of the left-hand side with respect to  $\Delta z$  is  $(z + \Delta z)^{-2}$ , while that of the right-hand side is  $\tau z^{-2}$ . When  $\Delta z \geq 0$ ,  $\tau z^{-2} \geq (z + \Delta z)^{-2}$ , so (9) holds. When  $\Delta z < 0$ , (9) holds as long as

$$\begin{aligned} \int_{\Delta z}^0 (z + \xi)^{-2} d\xi &\leq \int_{\Delta z}^0 \tau z^{-2} d\xi \\ [-(z + \xi)^{-1}]_{\Delta z}^0 &\leq \tau z^{-2} [\xi]_{\Delta z}^0 \\ \frac{1}{z + \Delta z} - \frac{1}{z} &\leq -\tau z^{-2} \Delta z \\ -\Delta z &\leq -\tau(1 + \frac{\Delta z}{z})\Delta z \\ 1 &\leq \tau(1 + \frac{\Delta z}{z}) \\ \frac{1}{\tau} - 1 &\leq \frac{\Delta z}{z} \end{aligned}$$

(In the third line from the bottom we multiply through by  $z(z + \Delta z) > 0$ , and in the second line from the bottom we divide through by  $-\Delta z > 0$ .)  $\square$

PROOF (LEMMA 6.3): Write  $\xi = S^{-1}\mathbf{1}$ , i.e.,  $\xi_i = 1/s_i$ . Write  $\mu = \frac{s^\top s}{n}$ . We have:

$$\begin{aligned} \|S^{-1}g\|^2 &= \|\beta\mu\xi - s\|^2 \\ &= \|\beta(\mu\xi - s) - (1 - \beta)s\|^2 \\ &= \|\beta(\mu\xi - s)\|^2 + \|(1 - \beta)s\|^2 \end{aligned}$$

since  $(\mu\xi - s)^\top s = n\mu - s^\top s = 0$ . The first term is a sum of squares, so is at least as large as any of its components:

$$\begin{aligned} \|S^{-1}g\|^2 &\geq \beta^2(\mu\frac{1}{s_0} - s_0)^2 + (1 - \beta)^2 s^\top s \\ &= \frac{\beta^2}{s_0^2}[(\mu - s_0^2)^2 + \frac{(1 - \beta)^2}{\beta^2} s_0^2 n\mu] \\ &= \frac{\beta^2}{s_0^2}[\mu^2 - \mu s_0^2 + s_0^4] \end{aligned}$$

since  $1 - \beta = \frac{\sqrt{n}}{n + \sqrt{n}}$ , so  $n\frac{(1 - \beta)^2}{\beta^2} = n\frac{(\sqrt{n})^2}{n^2} = 1$ . Finally, we can complete the square of  $(\frac{1}{2}\mu - s_0^2)$ , getting:

$$\begin{aligned} \|S^{-1}g\|^2 &\geq \frac{\beta^2}{s_0^2}[\frac{3}{4}\mu^2 + (\frac{1}{2}\mu - s_0^2)^2] \\ &\geq \frac{3}{4}\mu^2 \frac{\beta^2}{s_0^2} \end{aligned}$$

as desired.  $\square$

PROOF (LEMMA 7.1): Let  $D = (X + Y)^{-1}Y$ . Note that  $D$  is diagonal, with all elements strictly between 0

and 1 (the  $i$ th diagonal element is  $y_i/(x_i + y_i)$ ). Since  $\Phi^\top \Pi_0 = 0$ , we can rewrite the first line of (22) as:

$$G = \Phi^\top [\Phi U(D - I) - D] \Phi$$

The matrix  $\Phi U(I - D)$  is positive semidefinite: it has the same eigenvalues as its similarity transform

$$(I - D)^{1/2} \Phi U(I - D)^{1/2}$$

which is positive semidefinite since is of the form  $X^\top A X$  for real matrices  $A$  and  $X$  with  $A = \Phi U$  positive semidefinite. The matrix  $D$  is strictly positive definite, since it is diagonal with strictly positive diagonal elements. So, the sum  $\Phi U(I - D) + D$  is also strictly positive definite, as is  $\Phi^\top (\Phi U(I - D) + D) \Phi = -G$  since  $\Phi$  has full rank. So,  $G$  is invertible as claimed.  $\square$